

Consistency Models

Year: 2023 | Citations: 1367 | Authors: Yang Song, Prafulla Dhariwal, Mark Chen, I. Sutskever

Abstract

Diffusion models have significantly advanced the fields of image, audio, and video generation, but they depend on an iterative sampling process that causes slow generation. To overcome this limitation, we propose consistency models, a new family of models that generate high quality samples by directly mapping noise to data. They support fast one-step generation by design, while still allowing multistep sampling to trade compute for sample quality. They also support zero-shot data editing, such as image inpainting, colorization, and super-resolution, without requiring explicit training on these tasks. Consistency models can be trained either by distilling pre-trained diffusion models, or as standalone generative models altogether. Through extensive experiments, we demonstrate that they outperform existing distillation techniques for diffusion models in one- and few-step sampling, achieving the new state-of-the-art FID of 3.55 on CIFAR-10 and 6.20 on ImageNet 64x64 for one-step generation. When trained in isolation, consistency models become a new family of generative models that can outperform existing one-step, non-adversarial generative models on standard benchmarks such as CIFAR-10, ImageNet 64x64 and LSUN 256x256.